



# CONDUCT: An Expressive Conducting Gesture Dataset for Sound Control

Lei Chen, Sylvie Gibet, Camille Marteau

## ► To cite this version:

Lei Chen, Sylvie Gibet, Camille Marteau. CONDUCT: An Expressive Conducting Gesture Dataset for Sound Control. Language Resources and Evaluation Conference (LREC 2018), May 2018, Miyazaki, Japan. hal-01851389

**HAL Id: hal-01851389**

**<https://hal.science/hal-01851389>**

Submitted on 30 Jul 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# CONDUCT: An Expressive Conducting Gesture Dataset for Sound Control

Lei Chen, Sylvie Gibet, Camille Marteau

IRISA, Université Bretagne Sud

Vannes, France

{lei.chen, sylvie.gibet}@univ-ubs.fr, cam.marteau@hotmail.fr

## Abstract

Recent research in music-gesture relationship has paid more attention on the sound variations and its corresponding gesture expressiveness. In this study we are interested by gestures performed by orchestral conductors, with a focus on the expressive gestures made by the non dominant hand. We make the assumption that these gestures convey some meaning shared by most of conductors, and that they implicitly correspond to sound effects which can be encoded in musical scores. Following this hypothesis, we defined a collection of gestures for musical direction. These gestures are designed to correspond to well known functional effect on sounds, and they can be modulated to vary this effect by simply modifying one of their structural component (hand movement or hand shape). This paper presents the design of the gesture and sound sets and the protocol that has led to the database construction. The relevant musical excerpts and the related expressive gestures have been first defined by one expert musician. The gestures were then recorded through motion capture by two non experts who performed them along with recorded music. This database will serve as a basis for training gesture recognition system for live sound control and modulation.

**Keywords:** Corpus, sound-control gestures, expressive gesture, conducting

## 1. Introduction

The use of 3D mid-air gestures for controlling sound interactively has gained attention over this past decade (R. Aigner et al., 2012). One of the main difficulties to design such interfaces is the lack of guidelines to help designers in creating a set of gestures that can be significant for the given application and adopted by a large number of users.

The purpose of this study is to be able to control and modulate a live sound with expressive gestures. In live performances indeed the use of human movements to control both the visual and sound outputs have brought an intuitive and natural dimension to the interaction, leading to unprecedented performances, rich in body sensations and artistic creation. Among gestures controlling sound we are seeking a subset of meaningful and expressive gestures performed in the air, without any instrument. Furthermore, we expect the message conveyed by each gesture to correspond to an underlying desired function, and the quality encoded in the movement to correspond to an understandable expressive intent. The gestures should also be sufficiently different from each other, so that they may be automatically discriminable and recognizable. And finally, as in high-level languages, the structure of the produced gestures should allow to efficiently carry expressive variations by changing only a few gesture features such as kinematics, dynamics, geometry, or hand shape features.

Our search for a gestural language to drive sound applications naturally led us to consider conducting gestures performed by orchestral conductors. These gestures are of particular interest, since they gather most of the mentioned properties. Moreover, they are highly coded gestures which are potentially understandable by most of musicians around the world (Meier, 2009). Even if each conductor has his own style and defines his own set of gestures, we may find a subset of those gestures that share common meaning and

form.

Most descriptions of the conducting gestures are concerned by gestures executed by the dominant hand, i.e. beating gestures that indicate the structural and temporal organization of the musical piece (tempo, rhythm) and lead to precise, efficient, and unambiguous orders to the orchestral musicians. Other gestures performed by the non dominant hand are dedicated to show other aspects of the music execution and interpretation, among which variations in dynamics and intensity, musical phrasing or articulation, accentuation, entries and endings, sound quality and color, etc. These musical cues are sometimes noted as specific symbols or semantic terms in the score, but there is no agreement between musicians, musicologists or composers to precisely define the meaning of these additional notations on the scores. When the musical piece is performed, these indications can be translated into gestures that express the desired musical expression. For the musician, it comes down to interpreting, through his instrumental gesture, the nuance indicated on the score. For the conductor, the interpretation is understood more generally at the level of the ensemble of the musicians and the intention of the musical excerpt, and his gestures indicate in a slightly anticipatory manner the nuance that the musicians must achieve. We propose here to define a new data set of expressive gestures that are inspired by these conducting gestures performed by the non dominant hand. Our gesture selection has been partly guided by the joint work of professional conductors and linguists studying sign languages (Braem and Bräm, 2001). However, since we target non musicians in our interactive application, we chose intuitive, expressive, and easy to perform gestures, in order to efficiently control a musical performance.

In this paper, we present the design and construction of this new multimodal data set, called CONDUCT, composed of associated gestures and musical excerpts. To this end, we defined a set of interacting gestures that are likely to express and control the performance of music excerpts. We have identified four functional categories of gestures that corre-

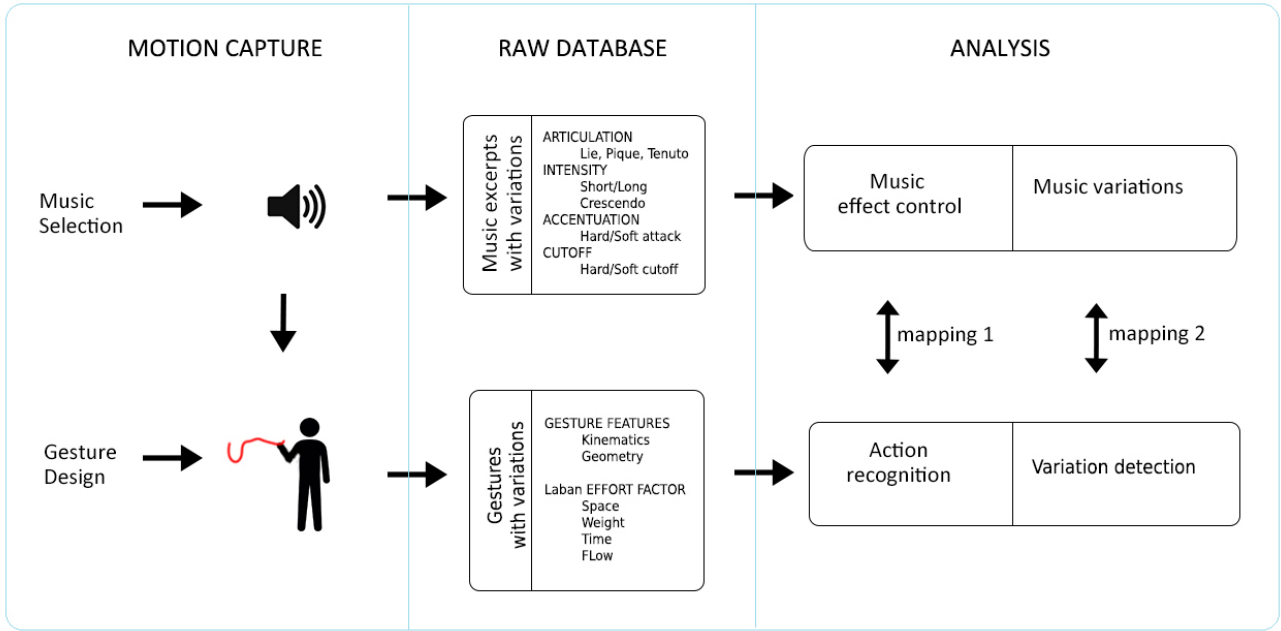


Figure 1: Work flow to collect the data and analyze them for the application needs.

spend to classically used musical effects. These categories relate to indications in Articulation, Intensity, Attack, or Cutoff. Within each category, we have characterized several expressive variations.

The paper is organized as follows: section 2 describes the main principles that have led to the design of the data set, both for the sound excerpts and for the design of associated gestures. Section 3 details the experimental setup, section 4 presents a first analysis of the data and section 5 concludes the paper and raises the main perspectives.

## 2. Overview of the Study

Figure 1 illustrates the way the data are collected (motion capture and sound recording data) and the future use of the database (recognition of actions and their variations). Concerning the experimental protocol, we first selected a set of appropriate musical excerpts that illustrate the main categories of sound effects (Articulation, Intensity, Accentuation, Cutoff) and their variations within each category. Our grammar of gestures follows the same protocol, by defining one specific gesture per category and variation.

Many researchers have designed experiments to record and analyze music-related gestures. Previous work has shown that gesture-sound correspondence is mutual, i.e. musical gestures help to understand the music, and, conversely, gestures can be used in turn to control the music. (Godøy et al., 2005) have studied *air-instrument* playing gestures where subjects were asked to observe real musician performances and imitate them as if they were playing the piano. In another study, they have asked the subjects to listen to the sound and draw induced sketches on a tablet (Godøy et al., 2006). Both studies contribute to the understanding of gesture-sound mapping. F. Bevilacqua et al. also proposed a procedure to learn the gesture-sound mapping where the users' gestures are recorded while they are listening to specific sounds (Bevilacqua et al., 2007) (Bevilac-

qua et al., 2011). On the basis of these theoretical studies, J. Françoise et al. have design several gesture-to-sound applications (Françoise et al., 2012) (Françoise et al., 2013). This work supports the approach we have adopted, by building the database from the music guiding the conductor rather than vice versa.

As illustrated in the right part of Figure 1 (Analysis), the aim of this gesture database is to serve as a basis for further gesture recognition for live sound control. Each recognized action will directly affect the sound effect (whether real or synthesized), and each detected gesture variation will correspond to a sound variation. In our approach the gestures are codified following sign language compositional structure, by identifying for each gesture the couple of components (hand movement, hand shape). With such a structure, it is possible to express the gesture variations as one modulation of either the hand movement (varying for example the form of the trajectory or the kinematic of the movement), or the hand shape (varying the shape of the hand). Previous work in sign recognition from video has demonstrated some success on isolated and continuous signing. Adding some linguistic knowledge about the composition of lexical signs considerably improves the performances of the recognition system (Dilsizian et al., 2014). Although these results are not achieved on motion capture data, they are promising for our training recognition system based on similar linguistic components.

## 3. Corpus Design

The purpose of this new data set was to be able to modulate a live sound with expressive gestures. From this hypothesis, the first questions that arose were: Which variables of the sound should we change? How much? In which categories should we consider such variations? Which variations should we take into account within the category? What qualifies as an expressive gesture? How is it related

to the sound variation? Instead of directly defining a set of non dominant hand gestures, as was done in (Braem and Bräm, 2001), we first listed different sound effects and grouped them in main functional categories. Among each category, we identified different variations that are commonly found in musical scores. We have then defined possible conducting-like gestures corresponding to these categories and variations, strongly inspired by the structural and iconic description of deaf sign languages.

### 3.1. Sound Categories and Variations

A musical score provides all the information needed to build an interpretation of the musical piece. Through this score, musicians (instrumentalists, singers) are able to read a musical piece and transform it in a gestural performance with musical sound. The challenge of the conductor is to have a global idea of the composer's musical intention, to imagine sounds and colors, and to read and understand all the scores of all the instruments. Among the information not contained in the temporal organization (tempo, rhythm) of the musical excerpt, we have identified four main categories: Articulation, Intensity, Attack, and Cutoff.

- The Articulation category is related to the phrasing of the musical discourse which is strongly dependent on the style of the piece. This expresses the way in which specific parts or notes in a piece are played within the musical phrasing, how they are linked and co-articulated, taking into account the timing and the quality of the musical sequencing. Among the techniques of articulation, we retained three of them: *Legato* (linked notes), *Staccato* (short and detached notes), and *Tenuto* (completely held and sustained). In our examples, we are aware that these terms and their meaning might differ according to the instrument and the musical context.
- The Intensity category, also called Dynamics in musicology, characterizes the music loudness. In our study we are interested in variations of intensity. These variations can be progressive (smooth) or abrupt with an increase or decrease in intensity. Four Intensity variations have been retained: *Long Crescendo*, *Long/Medium Decrescendo*, *Short Crescendo*, *Short Decrescendo*.
- The Attack category gathers different types of accents which are indicated in the score by different symbols but also by terms such as *sforzato* (*sfz*). In our study, we have identified two main discriminating attacks: *Hard hit*, *Soft Hit*.
- The Cutoff category expresses the way a musical phrase ends. We have retained two main variations within this last category: *Hard Cutoff*, *Soft Cutoff*.

These categories and variations have been retained for two kinds of musical excerpts:

- Orchestral classical music; some of the excerpts are taken from conducting scores (Meier, 2009);

- Two musical phrases with different variations played on a piano (one variation at a time, keeping the same tempo); we have chosen excerpts extracted from work of J. S. Bach: Prelude no. 1 in C Major, and Cantate Bwv 147.

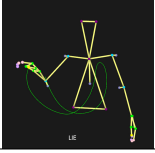
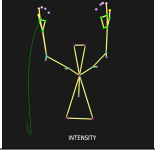
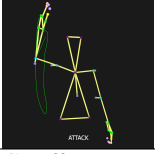
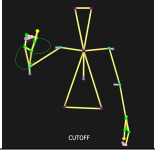
We also aim to add to this sound database two synthesized musical phrases similar to the above piano playing excerpts from J.S. Bach, along with corresponding generated variations for the four categories of actions.

### 3.2. Grammar of Gestures and their Modulation

For the conductor, the challenge is to interpret along the musical piece the sound effects and to give proper orders in terms of physical gestures to the musicians. Our aim was to find a set of intuitive and expressive gestures, in adequacy with the sound categories and variations. Beyond command and coverbal gestures (those that accompany the speech), several gestural languages have been developed for the exclusive expression by means of gestures. This is the case with gestures used by conductors or choirmasters, for whom it is impossible to express verbally. In addition, the need to teach the musical direction and to transmit it over time requires the definition of a form of codification of gestures and the definition of a written grammar. Another community that invented languages to express themselves and communicate with gestures is the one that practices the sign languages of the Deaf. Interestingly is the fact that there are strong similarities between these communities expressing themselves through gestures: they say by showing, they exploit the iconicity of gestures, and they mimic actions that make sense. Inspired by expressive conducting gestures we chose a subset of gestures that are shared by most of musicians around the world (Meier, 2009). In this study, we selected gestures composed of short actions, so that they can be used as isolated gestures, or repetitively as flows of actions. These gestures should implicitly contain the richness and the expressiveness of the conducting gestures, and they should be sufficiently simple to be codified and used in real-time through an automatic recognition process.

A novel grammar of gestures has therefore been defined in Table 1, whose structure is closely linked to sign language gestures. These gestures share indeed some common properties with conducting gestures. They are both visual-gestural languages: the produced information is multimodal and conveyed by hand-arm movements, facial expression and eye's gaze direction, and the perceivable information is interpreted by visual receptors, which means that the gestures have to be well articulated and visually comprehensible. Both of them have also a metaphoric function, i.e. they may fall into the category of describing objects (iconic gestures that describe the shape and the size of an object), or manipulating objects (like hitting or touching an object). As they may accurately express various meanings, they are characterized by a linguistic structure that can be described by a set of components also called phonological units in sign languages, including the location of the gesture, the movement, the hand shape, the wrist orientation and the facial expression. Therefore, as in high-level lan-

Table 1: List of gestures with their movement and handshape description.

| Gesture demo                                                                                      | with | Hand movement                                                                       | Handshape                                                                     |
|---------------------------------------------------------------------------------------------------|------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------|
| Articulation<br> |      | Phrasing gesture: Legato (sagittal plane); Staccato (horizontal); Tenuto (vertical) | Flat/Bent for Legato or Tenuto, O for Staccato                                |
| Intensity<br>    |      | Upward (crescendo) or downward (decrescendo)                                        | Alternating a Flat and a Bent handshape                                       |
| Attack<br>       |      | Metaphor of "hitting a hard object"                                                 | O for a hard attack and Fist for a soft attack                                |
| Cutoff<br>      |      | Metaphor of "taking an object out of the view"                                      | Flat handshape closes to a Pursed (soft cutoff) or O (hard cutoff) handshape. |

guages, the combination of these components allow to efficiently express various gestures and sequences of gestures. Modulating these gestures to express various meanings and nuances in the desired performances can be achieved by changing only one or few components (like the hand shape, or the kinematics and dynamics of the movement).

Our gesture database is composed of hand-arm gestures, and we only retain as structural components the hand shape and the hand movement. In sign languages the number and nature of hand shapes change according to the context of the corpus (around 60 basic hand shapes are identified in French Sign Language). This number is much more limited for conducting gestures. In our database we selected five basic hand shapes, as illustrated in Figure 2.

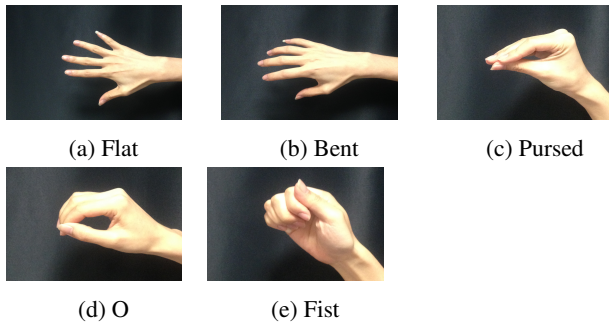


Figure 2: List of the five selected handshapes.

## 4. Experimental Protocol

To capture the conducting gestures, we used a Qualisys motion capture system which is a marker-based system composed of 12 high speed and high resolution cameras and one video camera. As these gestures mostly involve the hands and the upper body, we have fixed 8 markers on each hand, and 18 markers on the upper body, which counts for 34 markers in total. Figure 3 shows the MoCap marker setting. The MoCap frequency rate was set to 200 Hz to deal with the rapid movements of the hands. Two subjects of different musical levels (one non musician, and one musician highly graduated in instrument playing) participated in the experiment, both of them being right-handed. We call  $S_1$  and  $S_2$  these subjects.

Previously to the experiment, one expert musician (called  $E$ ), with extensive experience in orchestra, chamber music and instrumental violin practice, chose the musical excerpts, designed the corresponding gestures, and evaluated them qualitatively.

The MoCap recording experimentation can be described as follows: the musical excerpts were played as many times as necessary and each subject  $S_1$  and  $S_2$ , in two different sessions (each session lasted about four hours), was instructed to execute the corresponding gestures according to a specific musical effect. For each gesture, several variations were defined, induced by the musical excerpts. The users had to perform each gesture after a short training session, by visualizing some referent videos performed by the expert musician  $E$ . Each variation was expressed in several different musical excerpts. For example, for the orchestra music, we had about three excerpts per variation. Moreover, within the same musical excerpt, we could have different nuances of the same variation played at different times of the excerpt (for example several attacks, or several cutoffs). Each musical excerpt was played several times (up to 10 times).

For each subject, two recording sessions were considered, for two classes of classical music pieces:

- 30 classical orchestra music excerpts; these excerpts were validated by the expert musician  $E$ ;
- two piano excerpts from J.S. Bach played by a piano expert with the instructed variations.

All these excerpts covered the four sound control effects corresponding to the four categories (Articulation, Intensity, Attack, Cutoff), and for each category they covered the previously identified variations. Note that for the orchestra music excerpts, there were other variations related to the musical piece (style, articulation, rhythm, etc.) and the interpretation (tempo, loudness, etc.), whereas for the Bach piano pieces, the variations were more constrained. In the latter case, indeed, the same phrases were replayed several times with the instructed variations, but only one variation at a time, with the same tempo.

To sum up, we captured 50 gesture sequences from each participant, which corresponds to 50 musical excerpts. For each gesture sequence, each participant repeated the gesture at least 5 times. After pre-processing and manual cut-

ting, we obtained a conducting gesture data set of 1265 gesture units.

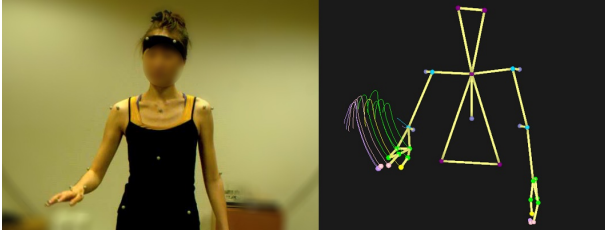


Figure 3: Mocap marker settings and marker-model reconstruction.

## 5. Analysis

Human movement can be represented by a regularly sampled sequence (also called time series) of  $T$  body postures. Each body posture corresponds in turn to a set of  $m$  joint positions  $\mathbf{x}(t) = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}(t)$ , where  $\mathbf{x}_k(t)$  (with  $1 \leq k \leq m$ ) is the 3D Cartesian position associated to the  $k_{th}$  joint and  $1 \leq t \leq T$ . Thus, a movement amounts to a matrix  $\mathbf{X} = \{\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(T)\}$  of dimensionality  $3 \times m \times T$ . The variation in movement category is illustrated

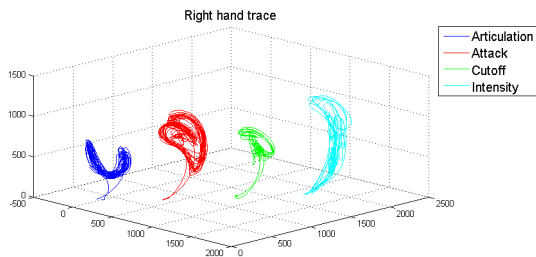


Figure 4: Right hand traces of one subject performing repetitively different gestures among the four categories: Articulation, Intensity, Attack, Cutoff.

in Figure 4. Four different right hand traces corresponding to the four categories (Articulation, Intensity, Attack, Cutoff) are repeated several times by one subject. We can see both the regularity of the traces within each gesture, and the variability of each gesture in shape and duration.

### 5.1. Feature Extraction and Vectorization

In order to characterize the expressive content of a given movement  $\mathbf{X}$ , we wished to consider a significant variety of descriptors, both for the recognition of gestures and for the detection of expressiveness, while ensuring that the selected quantities could be computed for several distinct movement categories and variations.

Based on a review of the motion descriptors traditionally used to characterize gestures (Karg et al., 2013), (Kapadia et al., 2013), (Larboulette and Gibet, 2015), we selected, (i) for the hand movement, kinematic features such as position, velocity, acceleration, and curvature, and, (ii) for the hand configuration, geometrical features measuring distances between the wrist and the extremity of the fingers (thumb, index finger, and middle fingers), as well as the volume

covered by the hand. The selected hand movement features have proven to be generic and sufficient to cover most variations of affect in movements (Carreno-Medrano et al., 2015). For example, for a right hand Cutoff gesture, illustrated by the trace of its captured data in Figure 5(a), we show the four kinematic features, i.e. the norm of the position, velocity, acceleration, and jerk (see Figure 5(b)). With regard to hand configuration features, they are specifically defined for the hand shapes of Figure 2.

In addition to these low-level descriptors, the Laban Effort descriptors (Maletic, 1987) are frequently used to characterize bodily expressiveness, especially in music-related gestures (Glowinski et al., 2011). Concerning conducting gestures, an experiment was also conducted to show how Laban Effort-Shape descriptors can help young conductors build expressive vocabulary and better recognize the level of expressiveness of different artistic disciplines (Jean, 2004). On the basis of these results, Laban Effort descriptors become a good candidate for characterizing the expressive variations of the conducting gestures. Focusing on the quality of motion in terms of dynamics, energy and expressiveness, these descriptors are defined by four sub-categories (Weight, Time, Space, and Flow). The Weight Effort refers to physical properties of the motion; the Time Effort represents the sense of urgency; the Space Effort defines the directness of the movement which is related to the attention to the surroundings; the Flow Effort defines the continuity of the movement. In summary, the chosen descriptors, whether they are kinematic, geometric, or inspired by Laban’s Effort components, are computed from the equations defined in (Larboulette and Gibet, 2015). They are listed in Table 2.

| Category           | Features                                                            |
|--------------------|---------------------------------------------------------------------|
| Kinematics         | Normalized velocity, acceleration, curvature, and jerk norms.       |
| Geometry and space | Displacements between hand joint (wrist) and each finger extremity. |
| Laban Effort       | Flow, Space, Time, and Weight.                                      |

Table 2: List of features used to characterize gesture and motion expressiveness.

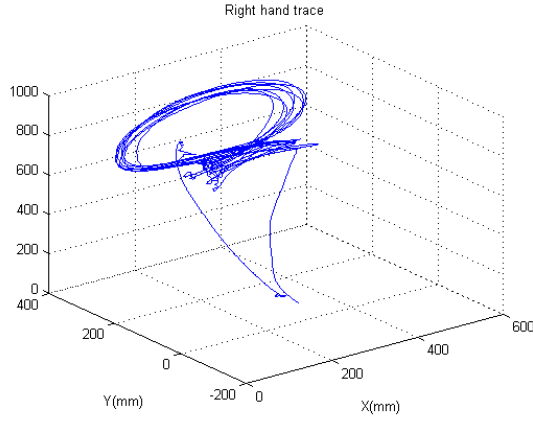
### 5.2. Recognition

For recognition purposes, we had to cut the gestures in elementary units, each cut gesture corresponding to a specific command and variation that will be used for controlling sound in live performances. The splitting has been achieved through a semi-automatic process, using the velocity information for all the hand traces. Moreover, to manage the variation of duration in our recognition process, we proceeded in two different ways, whether we consider off-line or on-line recognition.

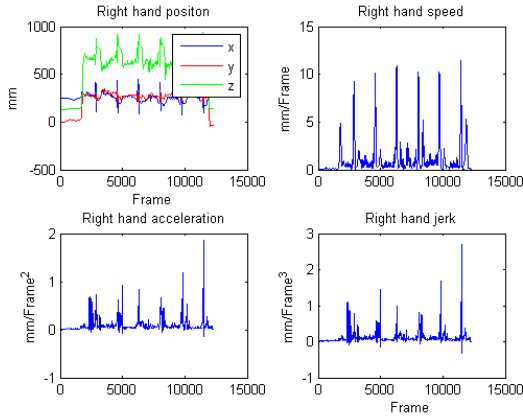
#### 5.2.1. Validation through Off-line Recognition

In the scope of validating our database, both for the gesture categories and the gesture variations within each category, we used an off-line classification approach applied on our gesture units. Two methods were considered to deal with the temporal variation of the gesture units. In the first





(a) 3D hand trace of the right hand.



(b) Feature sets.

Figure 5: Kinematic features of a Cutoff sequence.

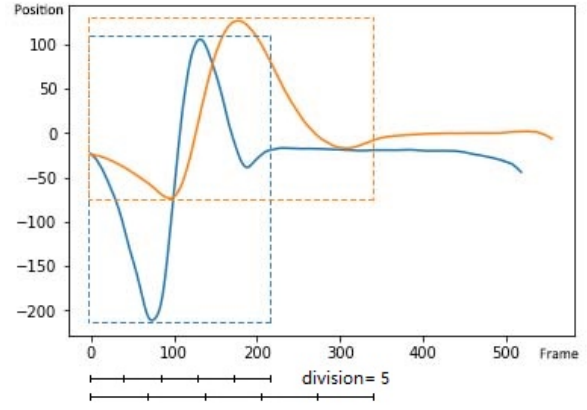
method, we divided each gesture unit into an equivalent number of segments (1, 5, 10 or 20 divisions), and computed a vector of normalized descriptors for each segment composing the gesture, using average and standard deviations of the kinematic and geometric features. In this way, each gesture is represented by a vector of the same size, which facilitates the recognition process. We plan to use in a second method elastic distances associated to Support Vector Machines methods.

### 5.2.2. On-line Recognition

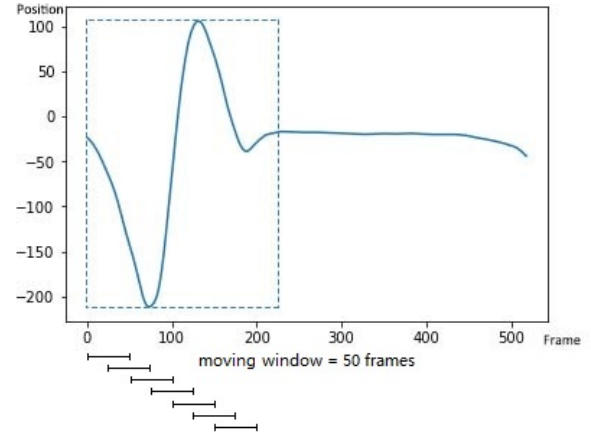
Secondly, for real-time recognition, we used a moving window and compute the average and standard deviation in this window. Figure 6 shows how the division method (part (a) of the Figure) or a moving window (part (b) of the Figure) works on a Cutoff gesture. These classification methods for off-line and on-line recognition are currently being implemented and tested.

## 6. Conclusion and Perspectives

In this article we have designed and recorded a novel database of expressive gestures induced by sound effects. The originality of this database lies in the fact that we have separated the functional components of the gestures from their expressive variation. This may lead to interesting applications where it is possible to control sound both from



(a) Equal division of two Cutoff gestures.



(b) Moving window on a Cutoff gesture.

Figure 6: Kinematic features of a Cutoff sequence.

the recognition of the action (gesture category) and from the recognition of the gesture expressiveness. In order to quantitatively validate the database, a recognition system is currently under development, using machine learning algorithms. This ongoing work currently focuses on two aspects: finding the best feature set for on-line classification, and adapting the algorithms for real-time recognition.

In the near future, we also intend to evaluate perceptually the database. The gestures will be evaluated by qualitative information related to Laban Effort components (Energy, Timing, Flow), and by multi-choice questionnaires using semantic terms, while the music excerpts will be evaluated through linguistic questionnaires.

## 7. Bibliographical References

- Bevilacqua, F., Guédy, F., Schnell, N., Fléty, E., and Leroy, N. (2007). Wireless sensor interface and gesture-follower for music pedagogy. In *Proceedings of the 7th international conference on New interfaces for musical expression*, pages 124–129. ACM.
- Bevilacqua, F., Schnell, N., Rasamimanana, N., Zamborlin, B., and Guédy, F. (2011). Online gesture analysis and control of audio processing. In *Musical Robots and Interactive Multimodal Systems*, pages 127–142. Springer.
- Braem, P. and Bräm, T. (2001). A pilot study of the ex-

- pressive gestures used by classical orchestra conductors. *Journal of the Conductor's Guild*, 22(1-2):14–29.
- Carreno-Medrano, P., Gibet, S., and Marteau, P.-F. (2015). End-effectors trajectories: An efficient low-dimensional characterization of affective-expressive body motions. In *2015 International Conference on Affective Computing and Intelligent Interaction, ACII 2015, Xi'an, China, September 21-24, 2015*, pages 435–441.
- Dilsizian, M., Yanovich, P., Wang, S., Neidle, C., and Metaxas, D. (2014). A new framework for sign language recognition based on 3d handshape identification and linguistic modeling. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, may. European Language Resources Association (ELRA).
- Françoise, J., Caramiaux, B., and Bevilacqua, F. (2012). A hierarchical approach for the design of gesture-to-sound mappings. In *9th Sound and Music Computing Conference*, pages 233–240.
- Françoise, J., Schnell, N., and Bevilacqua, F. (2013). A multimodal probabilistic model for gesture-based control of sound synthesis. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 705–708. ACM.
- Glowinski, D., Dael, N., Camurri, A., Volpe, G., Mortillaro, M., and Scherer, K. (2011). Toward a minimal representation of affective gestures. *IEEE Transactions on Affective Computing*, 2(2):106–118.
- Godøy, R. I., Haga, E., and Jensenius, A. R. (2005). Playing “air instruments”: mimicry of sound-producing gestures by novices and experts. In *International Gesture Workshop*, pages 256–267. Springer.
- Godøy, R. I., Haga, E., and Jensenius, A. R. (2006). Exploring music-related gestures by sound-tracing: A preliminary study. In *Proceedings of the COST287-ConGAS 2nd International Symposium on Gesture Interfaces for Multimedia Systems (GIMS2006)*, pages 27–33.
- Jean, N. E. (2004). *The effect of Laban Effort-Shape instruction on young conductors' perception of expressiveness across arts disciplines*. Ph.D. thesis, University of Minnesota.
- Kapadia, M., Chiang, I., Thomas, T., Badler, N. I., and Jr., J. T. K. (2013). Efficient motion retrieval in large motion databases. In *Symposium on Interactive 3D Graphics and Games, I3D '13, Orlando, FL, USA, March 22-24, 2013*, pages 19–28.
- Karg, M., Samadani, A., Gorbet, R., Kühnlenz, K., Hoey, J., and Kulic, D. (2013). Body movements for affective expression: A survey of automatic recognition and generation. *IEEE Transactions on Affective Computing*, 4(4):341–359.
- Larboulette, C. and Gibet, S. (2015). A review of computable expressive descriptors of human motion. In *Proceedings of the 2nd International MoCo Workshop*, pages 21–28. ACM.
- Maletic, V. (1987). *Body-space-expression: The development of Rudolf Laban's movement and dance concepts*, volume 75. Walter de Gruyter.
- Meier, G. (2009). *The Score, the Orchestra, and the Conductor*. Oxford University Press.
- R. Aigner, D., Wigdor, H., Benko, M., Haller, D., Lindbauer, A., Ion, S., Zhao, J., and Koh, T. (2012). Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for hci. Technical report, November.